Semaine d'Etude Maths-Entreprises 8 : Optimisation du voyage d'un commercial en environnement de rentabilité hétérogène

François Dayrens, Isai Lankoande, Mathilde Legrand, Nga Nguyen, Anthony Preux, Yacouba Samoura

Semaine d'Etude Maths-Entreprises

Optimisation du voyage d'un commercial en environnement de rentabilité hétérogène.

Responsable projet : Delsol Laurent

Présenté par :

Dayrens François
Lankoande Isai
Legrand Mathilde
Nguyen Nga
Preux Anthony
Samoura Yacouba

Remerciements

Nous remercions tous les organisateurs de cette huitième Semaine d'Etude Maths-Entreprises, en particulier le laboratoire du MAPMO pour son accueil et bien entendu AMIES pour le financement. Le sujet présenté ici nous a été proposé par l'entreprise ARTICQUE par l'intermédiaire de Jérôme Barthélémy qui a bien pris le temps de nous en exposer les tenants et aboutissants. Nous remercions naturellement Laurent Delsol qui a suivi notre groupe durant toute la semaine mais aussi Romain Abraham et Mounir Haddou pour leurs conseils scientifiques.

1 Présentation

ARTICQUE est une entreprise spécialisée dans la cartographie. Dans ce cadre, elle est amenée à vendre des cartes dans toute la France au travers des stations services situées sur les autoroutes. Ces stations n'ont pas la possibilité de gérer elles-mêmes les stocks, elles ne sont qu'un dépôt de vente. Les commerciaux d'ARTICQUE ont donc la responsabilité de passer sur les sites afin d'établir les commandes de cartes.

La répartition des sites entre les différents commerciaux (visiteurs) a été faite en amont de cette étude. Elle permet, grâce à sa configuration géographique, de dissocier le travail des différents visiteurs. En effet, chaque site est attribué à un seul et même visiteur (voir figure 1). Ainsi l'étude réalisée ici se concentre sur la stratégie de visites d'un seul visiteur sur sa zone.

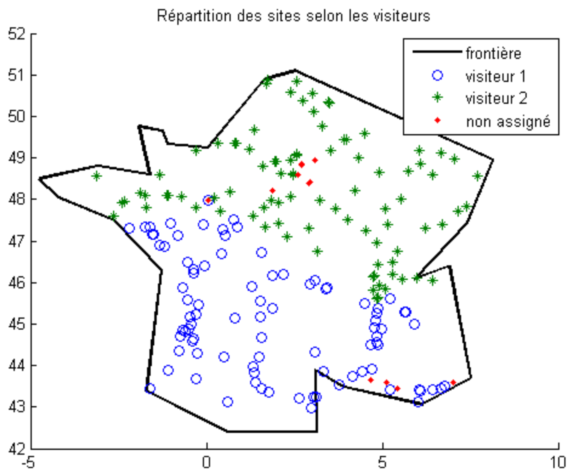

FIGURE 1 – Sites de ventes

Remarque 1 *Les sites non assignés (en rouge sur la figure 1) seront redistribués au visiteur le plus probable géographiquement parlant.*

Un visiteur est actuellement libre d'appliquer sa propre stratégie dans la visite des sites. Il lui est par exemple possible de visiter tous les sites dans un ordre précis, sur le simple critère géographique. Cependant, les sites ont des débits d'écoulement différents. Par conséquent si le rythme de passage est le même pour toutes les stations, certaines seront en rupture de stock quand d'autres auront encore suffisamment de marchandises. De par ce critère d'intérêt d'un site par rapport à un autre, ce travail se distingue du classique problème du voyageur de commerce ([3], [2]). Il n'est pas nécessaire de passer par tous les sites mais il faut déterminer quels sites doivent être visités en priorité. Le projet SEME a donc consisté en une proposition d'une tactique de visites plus élaborée pour les commerciaux prenant en compte le chiffre d'affaire des points de vente.

Avant d'aborder pleinement le modèle mathématique décrivant le trajet d'un visiteur, voyons plus en détail le principe de déplacement d'un commercial. Celui-ci a une tournée à effectuer sur un temps donné, par exemple un jour ou une semaine. Il connaît d'avance certaines contraintes : son point de départ et celui d'arrivée. A partir de ces critères, le visiteur doit établir le programme des visites durant toute sa tournée.

La résolution du problème a été effectuée en gardant comme fil conducteur le cas particulier décrit par l'entreprise ARTICQUE. Cependant, elle ne s'y restreint pas. Un certain nombre de paramètres libres apparaîtront permettant d'appliquer l'algorithme à différents jeux de données. Le seul impératif, évoqué plus haut, consiste à n'avoir qu'un seul visiteur possédant ses propres sites qu'il ne partage avec personne. Dans la suite de ce rapport, nous présenterons tout d'abord une modélisation du problème permettant d'obtenir une feuille de route indiquant les sites à visiter durant la tournée. Dans la troisième et dernière partie, une résolution allégée sera décrite et illustrée par une simulation sur trois semaines.

2 Modélisation du problème

On décrit ici la mise en équation qui nous permettra de résoudre notre problème, c'est-à-dire de donner une liste des sites à aller visiter par le commercial.

2.1 Intérêt des sites

On note $(S_i)_{i\in I}$ l'ensemble des sites gérés par le visiteur (l'indice i est une numérotation arbitraire des sites) auquel on ajoute deux sites fictifs S_{i_D} et S_{i_A} qui correspondent au point de départ et au point d'arrivée de la tournée. Ces points ne sont pas forcément des sites, le voyageur peut commencer sa semaine en partant de son domicile et la finir par un rapport au siège social de la firme. A part les sites fictifs, on définit un potentiel de vente :

$$PV_i = f_i\ \Delta T_i$$

avec f_i une fonction fréquence (positive) et ΔT_i la durée (en jour ou en semaine) depuis la dernière visite.

La fonction fréquence sera décrite plus loin. L'idée est de dire que plus f_i est grande, plus le site fait des ventes et donc plus il faudra y retourner rapidement. Ce potentiel de vente nous permet donc de quantifier la nécessité de rendre visite à un site par rapport aux autres pondéré par le temps écoulé depuis la dernière visite. A ce titre, il n'est pas nécessaire de passer par les sites S_{i_D} et S_{i_A} en cours de tournée, ce qui s'indique par $PV_{i_D} = PV_{i_A} = 0$.

2.2 La fonction fréquence

La fonction fréquence détient le rôle clef dans l'évaluation du potentiel de vente. C'est elle qui doit juger de la valeur comparative des sites. C'est naturellement sur elle qu'on laisse le plus de liberté afin de faire coller le modèle à la réalité. Son expression est donc la plus difficile à déterminer. On peut imaginer de nombreux paramètres pour une telle fonction :

$$f_i = f(CA_i, ES_i, P, G_i).$$

L'indice i se rapporte au site S_i. CA_i est le chiffre d'affaire du site, plus il est élevé et plus le site est intéressant. ES_i est l'état des stocks du site, on peut construire une fonction qui explose lorsque les stocks sont dangereusement bas afin de traduire la prévention d'une rupture de stock. G_i est la position géographique du site à mettre en relation avec la période de l'année P : les stations de ski sont plus importantes en hiver, de même pour les stations balnéaires en été.

Tout ces paramètres ne sont pas forcément indépendants. En effet, un site qui réalise un gros chiffre d'affaire écoule rapidement ses stocks. Ainsi on peut exprimer ES_i en fonction de CA_i et de ΔT_i. Par exemple :

$$ES_i = SM_i - CA_i^* \Delta T_i$$

avec SM_i le stock maximal possible du site et CA_i^* une "vitesse de vente" hebdomadaire ou journalière indexée sur le chiffre d'affaire. f_i peut ainsi dépendre de ΔT_i aussi.

Il est certainement possible d'exprimer une fonction fréquence extrêmement détaillée qui prend en compte exhaustivement plusieurs produits selon leur importance, le fait que certains peuvent tomber en rupture de stock et d'autre non *etc.*

Le manque de connaissance de la politique choisie par les visiteurs et le manque de données fournies par ARTICQUE ne nous ont pas permis de construire une fonction très compliquée. Avec seulement la situation géographique et le chiffre d'affaire annuel de chaque site, nous avons décidé de prendre

$$f_i = f(CA_i^{annuel}).$$

où f est une fonction croissante. Cette forme encode déjà bien le comportement des sites. En effet, à chiffre d'affaire élevé, la fréquence est élevée ce qui fait qu'on prévient la rupture de stock. La fonction f permet ainsi de calibrer l'importance des gros sites (gros chiffres d'affaire) par rapport à celle des petits (par exemple en déterminant à quel seuil va-t-on visiter deux fois plus souvent un gros site qu'un petit). Dans les simulations qui suivront, nous avons pris $f(CA_i) = CA_i$ pour faire un cas vraiment simple.

2.3 Temps de parcours

Quelque soit l'intérêt d'un site, si on ne peut pas l'atteindre, il est inutile de le sélectionner. Ainsi une contrainte importante du problème est le temps mis par le voyageur pour faire sa tournée. On différencie deux temps différents : le premier est le temps de traitement de chaque site (évalué à deux heures d'après ARTICQUE) et le deuxième est le temps de trajet pour aller d'un site au suivant. Ce temps de cheminement, noté T_c, est compliqué à calculer et est un problème d'optimisation à part entière. On décrit ce temps sans trop entrer dans les détails.

En une semaine, le voyageur peut visiter au maximum une quinzaine de sites (à deux heures de traitement par site, cela fait déjà 30 heures sur les 35 d'une semaine classique!). On note $(S_i)_{i\in I_0}$ l'ensemble des sites pour lesquels on veut calculer le temps de trajet et d_{ij} la distance entre deux sites S_i et S_j (durée du trajet entre ces deux sites). Le temps de parcours associé à cet ensemble de site est le problème du *facteur chinois* ([3], [2]), c'est-à-dire trouver le chemin qui permet de rejoindre tous les sites (en repassant éventuellement par des sites déjà visités) avec la durée minimale. Une fois ce chemin trouvé, le temps de parcours est facile à calculer.

Théorème 1 *Soit $S_{i_1}, S_{i_2}, \dots, S_{i_N}$ un ensemble de sites à aller voir et (d_{ij}) les distances entre deux de ces sites. Il existe un chemin de coût minimal qui parcours tous ces sites (avec éventuellement des répétitions) :*

$$S_{i_{k_1}} \to S_{i_{k_2}} \to \cdots \to S_{i_{k_n}}$$

et le temps de parcours associé est donné par

$$T_c = \sum_{j=1}^{n-1} d_{i_{k_j} i_{k_{j+1}}}.$$

On a donc une fonction "boîte noire" qui, à un sous ensemble de sites I_0 nous renvoie le temps de parcours de ces sites $T_c(I_0)$.

Par cette méthode, $T_c(I_0)$ est calculable mais beaucoup trop coûteux pour un ordinateur étant donné qu'il ne s'agit que d'une fonction auxiliaire qui sera appelée de nombreuses fois dans un algorithme d'optimisation. On présentera plus loin une fonction plus simple et plus grossière pour évaluer le temps de parcours.

A ce temps de trajet entre les sites, il faut ajouter le temps de traitement de chaque site. Dans notre cas, c'est le même pour tous les sites, mais pour la théorie ce n'est pas plus compliqué de les supposer différents. A chaque site S_i on associe un temps de traitement τ_i et le temps de traitement de l'ensemble des sites $(S_i)_{i \in I_0}$ est alors donné par

$$\sum_{i \in I_0} \tau_i.$$

Ceci nous permet de calculer le temps de la tournée $(S_i)_{i \in I_0}$.

Définition 1 *Soit I_0 un sous ensemble de I correspondant à la tournée de la semaine ou de la journée. Le temps de la tournée est donné par*

$$T(I_0) = T_c(I_0) + \sum_{i \in I_0} \tau_i.$$

2.4 Modélisation

Ici commence vraiment la partie mathématique. Soit X un vecteur colonne de taille $N = \text{card}(I)$, $X_i = 0$ signifie qu'on ne va pas visiter le site S_i et $X_i = 1$ signifie qu'on va le visiter. Soit τ le vecteur de la même taille que X tel que τ_i soit le temps de traitement du site S_i. Le temps de trajet calculé précédemment se traduit ici par $T_c(X)$ qui est le temps de trajet de l'ensemble des sites S_i tel que $X_i = 1$. Et enfin, soient i_D et i_A les points de départ et d'arrivée de la tournée et t la durée de cette tournée (typiquement $t = 8$ heures pour une journée et $t = 35$ heures pour une semaine).

La feuille de route du voyageur de commerce est donnée par la résolution du problème d'optimisation suivant :

$$\max_X \sum_{i \in I} X_i \ PV_i$$

sous les contraintes

$$X_{i_D} = X_{i_A} = 1, \quad T_c(X) + \sum_{i \in I} X_i \tau_i \leqslant t.$$

La contrainte sur le point de départ et/ou sur celui d'arrivée peut être enlevée. De même le temps de traitement de ces deux sites fictifs peut soit être nul, soit correspondre à un temps de travail "hors site" que le commercial souhaite comptabiliser durant son voyage. D'une semaine sur l'autre (ou d'un jour sur l'autre) le PV_i des sites est mis à jour : celui des sites visités tombe à zéro car le ΔT_i est réinitialisé à 0 et celui des sites non visités augmente. Ce processus permet de voir que d'une tournée à la suivante, la feuille de route est différente.

Remarque 2 *Lors du calcul des PV_i, on peut tenir compte des sites voisins en calculant un autre potentiel de vente à partir des potentiels "internes" de chaque site :*

$$PV_i^* = PV_i + \frac{1}{2} \sum_{S_j \, voisin} PV_j$$

ou

$$PV_i^* = PV_i + \sum_{j \in I, i \neq j} \frac{PV_j}{d_{ij}}.$$

Ceux-ci ne sont que deux exemples de potentiel possibles, le facteur $\frac{1}{2}$ ou le terme $\frac{1}{d_{ij}}$ indique une dépréciation de l'intérêt du site voisin.
Nous avons choisi de ne pas utiliser cet autre potentiel car le regain d'intérêt induit par le "groupement" de sites plus ou moins importants est déjà pris en compte dans la contrainte du temps de la tournée. En effet, si plusieurs sites sont intéressants dans une même zone alors on gagne beaucoup au niveau du terme T_c et il reste donc du temps permettant de voir plus de sites.

2.5 Mise à jour dynamique

Le chiffre d'affaire est le point clé de l'intérêt d'un site. Ce chiffre d'affaire peut être renseigné grâce aux ventes de l'année précédente et resté fixé pour la nouvelle année. Cependant, il est raisonnable d'envisager sa variation. Avec le modèle simple considéré dans la section 2.1

$$PV_i = f(CA_i)\Delta T_i,$$

il est possible de tenir compte de l'évolution de la capacité de chaque site. Concrètement, si un site a fait un chiffre d'affaire faible lors du calcul de son PV et qu'on suppose qu'il est capable de faire bien mieux alors sa progression est bornée par le fait qu'on ne va pas le voir souvent (il est systématiquement en rupture de stock) et donc sa rentabilité en est très amoindrie. L'idée est, par exemple pour des tournées d'une semaine, de calculer un chiffre d'affaire hebdomadaire a priori pour chaque site et de mettre à jour ce CA à chaque visite selon les constats du commercial. Si les stocks sont vides alors la prochaine fois il faudra revenir plus tôt, pour cela on augmente son CA. De la même manière, si les stocks restants sont très élevés alors il n'est pas nécessaire de revenir aussi souvent, donc on baisse son CA. Avec cette idée, le CA est plus un indicateur de la capacité du site à écouler les stocks qu'un réel chiffre d'affaire.

Voici un exemple de dynamisation du CA qui accorde une place importante aux ruptures de stock.

$$CA_i^{nouveau} = \lambda CA_i^{ancien} + (1 - \lambda)\frac{v_i}{\delta t_i}$$

avec v_i les ventes réalisées depuis la dernière visite, $\lambda \in [0, 1]$ indique l'inertie souhaitée pour l'évolution du CA. δt_i correspond au temps depuis la dernière visite sachant que

$$\delta t_i = \begin{cases} \Delta T_i & \text{s'il reste du stock} \\ \Delta T_i - r_i & \text{si le site est en rupture depuis } r_i \text{ semaines} \\ \Delta T_i / 2 & \text{si on ne sait pas depuis combien de temps le site est en rupture.} \end{cases}$$

2.6 Indicateur de sectorisation

Grâce au potentiel de vente, introduit pour optimiser la tournée du commercial, il est possible d'avoir deux indicateurs sur la qualité de la sectorisation. En effet à la n-ième semaine, on peut calculer deux grandeurs :

$$E_n = \frac{1}{N} \sum_{i \in I} PV_i^n$$

et

$$M_n = \max_{i \in I} PV_i^n.$$

On rappelle que N =card(I) et PV_i^n est le potentiel de vente du site S_i lors de la n-ième semaine. Si E_n a tendance à augmenter alors cela signifie que le commercial n'a pas le temps de visiter l'ensemble des sites pour palier à la demande sur l'ensemble de son secteur. De même si E_n diminue alors il pourrait se charger de plus de sites. La grandeur M_n mesure plus l'importance des gros sites sur son secteur, si elle augmente alors il y a un risque que les petits sites de son secteur soient laissés à l'abandon car peu rentables à coté des gros sites.

En comparant les E_n et M_n des différents commerciaux, on peut adapter la sectorisation de sorte à avoir un meilleur équilibre.

3 Applications numériques

3.1 Difficultés rencontrées

Même si la modélisation mathématique utilise des outils théoriques très simples pour trouver la solution exacte, elle est inutilisable en pratique car elle est non constructive. Le choix d'un algorithme d'optimisation sous contraintes ainsi que son efficacité dépendent fortement de la fonction à minimiser et des fonctions contraintes (linéaire, quadratique, convexe, différentiable, polynomiale...) et de la forme de la solution (vecteur à coordonnées entières ou quelconques).La complexité est un domaine à part entière de l'informatique qui étudie, entre autre, le temps d'éxécution d'un algorithme en fonction de la taille des données entrantes (dans notre cas, le nombre de site et la distance entre chacun d'eux). Les problèmes NP, comme le calcul du temps de parcours dans notre étude, sont les plus délicats à résoudre par l'informatique. Nous renvoyons à [5] pour des explications avancées sur l'agorithmie et les problèmes "NP".

En ce qui nous concerne, nous avons décidé de simplifier le temps de parcours T_c afin de pouvoir calculer des solutions sans avoir recours à des calculateurs puissants (la finalité du travail étant de pouvoir effectuer les calculs en un temps raisonnable sur des ordinateurs portables). Il y existe plusieurs possibilités de simplification. La meilleure serait de la rendre linéaire ou quadratique, ce qui simplifierait grandement le calcul de la solution, même si elle risque de ne plus être la meilleure. Voilà la solution que nous avons apportée à ce problème : la fonction T fut remplacée par la fonction quadratique suivante :

$$\tilde{T}(I_0) = \sum_{i,j \in I_0} D_{ij} X_i X_j$$

où D_{ij} représente la durée de parcours entre le site i et le site j (le lien avec l'expression précédente est détaillé en section 3.2).

Cette fonction a de bonnes qualités :

1. Elle permet de pénaliser grandement un site trop loin par rapport à un regroupement de sites dont le potentiel de vente serait en moyenne un peu plus bas mais qui permettrait d'en visiter plusieurs très rapidement.
2. On peut imposer le passage sur un site, il suffit d'augmenter le potentiel de vente de celui-ci de tel sorte qu'il soit plus important que tous les autres réunis.

Nous allons pouvoir constater que, malgré cette simplification, l'algorithme donne de bons résultats sur le test que nous lui avons proposé. Finalement on peut remplacer le problème par le suivant :

Trouver λ maximal tel que :

$$\begin{cases} \min_{I_0 \subset I} \sum_{i,j \in I_0} D_{ij} X_i X_j < d \\ \sum_{i \in I_0} PV_i > \lambda \end{cases} \tag{1}$$

Ce qui devient une minimisation d'une fonction quadratique sous contrainte linéaire, problème de complexité polynomial (voir section 3.2 pour plus de détails). Le scalaire d est une variable ajustable selon le commercial, si celui-ci est rapide sur les sites, il peut alors visiter plus de sites ou des sites plus loin et donc avoir un d plus grand que les autres commerciaux.

3.2 Optimisation numérique et simulations

Dans cette partie, nous expliquons en premier lieu comment nous avons résolu, numériquement, le problème d'optimisation posé précédemment puis proposons une illustration des résultats (affichage Matlab). Le problème consiste à trouver λ tel que :

$$\begin{cases} \min_{I_0 \subset I} \sum_{i,j \in I_0} D_{ij} X_i X_j < d \\ \\ \sum_{i \in I_0} PV_i > \lambda \end{cases}$$

où d est un paramètre ajustable par le visiteur.
En réalité, ce problème est une contraction de deux problèmes, comme expliqué en section 2.3. Pourquoi une contraction de deux problèmes ? C'est tout simplement parce que le problème précédent peut être réécrit comme

$$\begin{cases} \max \lambda \\ \text{sous les contraintes :} \quad \min_{I_0 \subset I} \sum_{i,j \in I_0} D_{ij} X_i X_j < d \\ \\ \qquad\qquad \sum_{i \in I_0} PV_i > \lambda \end{cases}$$

On voit clairement que pour chaque "tour" du problème d'optimisation principal

$$\begin{cases} \max \lambda \\ \\ \text{sous des contraintes} \end{cases}$$

il faut résoudre un problème d'optimisation intermédiaire

$$\min_{I_0 \subset I} \sum_{i,j \in I_0} D_{ij} X_i X_j < d$$

pour d fixé.

En section 2.4, nous avons donné le problème "simplifié", celui qui nous évite de passer par un problème intermédiaire. Il stipulait que nous devions résoudre

$$\begin{cases} \max \quad \sum_{i \in I} PV_i X_i \\ \\ \text{s.c.} \quad T_c(X) + \sum_{i \in I} \tau_i X_i \leq t \end{cases}$$

où t est une donnée fixée d'avance (8 heures pour le jour, 35 heures pour la semaine). Il en ressortait que bien que le problème d'optimisation intermédiaire ait été levé, un autre problème apparaissait : celui de la complexité du calcul de T_c. Mais il est facile de remarquer la chose suivante :

Imposer une contrainte de temps (8 heures par jour ou 35 heures par semaine ou toute autre valeur par unité de temps choisie) revient à imposer une contrainte de distance, vu que la distance et le temps sont deux paramètres liés.

Ce qui transforme la contrainte

$$T_c(X) + \sum_{i \in I} \tau_i X_i \leq t$$

en la contrainte quadratique

$$\sum_{i,j \in I_0} D_{ij} X_i X_j \leq d$$

avec d fixé par l'entreprise ARTICQUE. Ici, D_{ij} ne représente plus le temps entre les sites i et j mais plutôt la distance entre ces deux sites.

Remarque 3 *Ce passage à une distance plutôt qu'à un temps n'est pas nécessaire. Il s'est simplement trouvé que les données fournies par* ARTICQUE *apportaient des informations géographiques et que nous n'avons pas eu le temps de convertir cela en un temps de trajet. C'est pourquoi nous introduisons cette conversion afin d'être en accord avec l'algorithme implémenté.*

Enfin, est-on tenté de dire, on peut récapituler notre problème d'optimisation en la chose suivante :

$$\left\{ \begin{array}{ll} \max & \displaystyle\sum_{i=1}^{N} PV_i X_i \\ \\ \text{s.c.} & \displaystyle\sum_{i,j=1}^{N} D_{ij} X_i X_j \leq d \end{array} \right. \tag{2}$$

avec N le nombre de sites à visiter et d la limite sur la distance.

Dans le jeu de données que l'entreprise nous a fourni, nous avons trouvé 106 sites à visiter ($N = 106$). Dans ce jeu de données, on retrouve aussi les chiffres d'affaire annuels (ce qui nous permet de calculer les chiffres d'affaire hebdomadaires) et des coordonnées géographiques des sites (latitudes et longitudes). Pour les sites i, j avec leurs coordonnées, latitudes L_i, L_j et longitudes l_i, l_j, nous calculons la distance par la formule :

$$D_{ij} = \sqrt{(L_i - L_j)^2 + (l_i - l_j)^2}$$

Remarque 4 *L'utilisation des coordonnées géographiques plutôt que de réelles distances ne permet pas d'utiliser une distance d limite très explicite. Nous avons pu observer que $d = 300$ permettait d'obtenir un nombre raisonnable de sites. Ainsi dans les simulations qui suivent c'est cette valeur $d = 300$ qui a été introduite.*

Les chiffres d'affaire annuels donnés ont été divisés par 52 pour avoir les chiffres d'affaire hebdomadaires. En section 2.5, nous prenons comme potentiel de vente $PV_i = f(CA_i)\Delta T_i$. En section 2.2, la fonction de fréquence a été simplifiée avec $f(CA_i) = CA_i$. Comme les calculs se font par semaine (unité de temps = 1 semaine), $\Delta T_i = 1$. Ainsi, $PV_i = CA_i$ tout simplement pour l'initialisation.

La variable X est un vecteur de binaires. Le critère est linéaire et la contrainte quadratique (voir [4]). Donc, nous nous retrouvons avec un problème d'optimisation entière non linéaire ou Integer Non-Linear Programming (en anglais INLP [1]). Pour sa résolution numérique, nous avons utilisé le langage de programmation mathématique AMPL [1]. Avec ce dernier, nous avons créé 3 fichiers :

- le fichier `.mod` contenant le model du problème (2).
- le fichier `.dat` contenant les données du problème. Par données du problème, comprenez les valeurs du nombre de sites N, des latitudes L, des longitudes l, des chiffres d'affaire annuels CA et de la limite sur la distance d.
- le fichier `.ampl` contenant les commandes de résolution en l'occurrence le solver utilisé. Parlant de solver, il faut savoir que compte tenu du fait qu'on a un type particulier de Integer Programming, nous avons utilisé CPLEX comme solver.

La résolution du problème 2 fournit le vecteur binaire X, donnant ainsi les informations sur les sites à visiter.

Voyons sur une simulation sur trois semaines les résultats obtenus grâce à cette stratégie. Nous avons initialisé cette illustration à l'aide du chiffre d'affaire hebdomadaire moyen pour tous les articles, seule donnée en notre possession. Et nous avons considéré que tous les ΔT_i étaient égaux. La figure 2 montre le potentiel de vente ainsi calculé pour le commercial du sud de la France, désigné pour tester l'algorithme. La figure 3 nous donne les sites choisis par l'algorithme de calcul, on remarque qu'il décide d'aller visiter le point à fort potentiel de vente situé au centre ouest malgré le fait qu'il soit éloigné d'autres points intéressants. Cela montre que l'algorithme peut tout de même retenir des sites éloignés s'ils représentent une forte probabilité de bénéfices. On a continué l'algorithme sur quelques semaines pour voir les résultats obtenus, comme nous n'allons pas réellement sur les sites, nous ne pouvons pas mettre à jour le potentiel de vente comme il l'a été décrit dans la partie précédente. Nous avons donc décider de faire comme si les sites vendaient par semaine exactement le nombre d'articles équivalent à leurs chiffres d'affaire moyens hebdomadaires. Dans ce cadre, les résultats obtenus sont cohérents avec nos attentes, de plus nous obtenons les résultats instantanément grâce au problème simplifié. Le seul facteur d'intérêt variable devient donc le temps écoulé depuis la dernière visite. A chaque fois qu'un site est visité, son ΔT_i est réinialisé et son potentiel de vente diminue ainsi fortement. Autrement, le ΔT_i est incrémenté de 1. En raison de cette variation des potentiels de vente, il est peu probable de retourner deux fois en deux semaines consécutives sur le même site sauf si celui-ci a un chiffre d'affaire bien supérieur aux autres. On remarque qu'en semaine 3 (voir figure 7) des sites visités la première semaine sont revisités, ceci est du au fait que quelques sites ont un énorme potentiel de vente par rapport aux autres et qu'il est donc important de les revisiter rapidement.

1. http://ampl.com/, http://en.wikipedia.org/wiki/AMPL

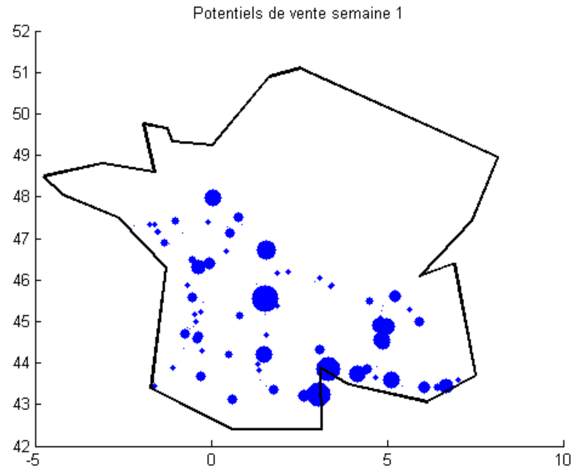

FIGURE 2 – Le potentiel de vente lors de la première semaine.

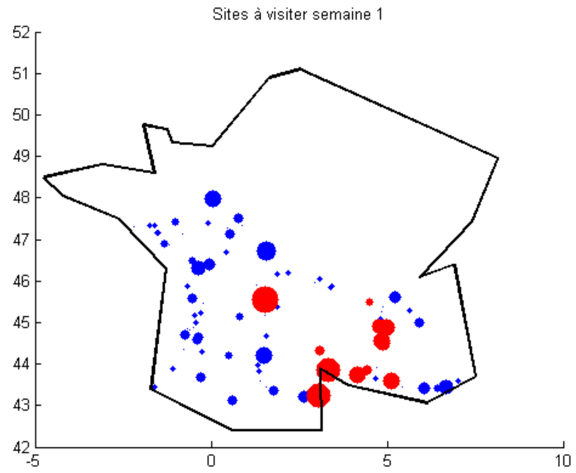

FIGURE 3 – Les sites choisis (en rouge) pour la première semaine.

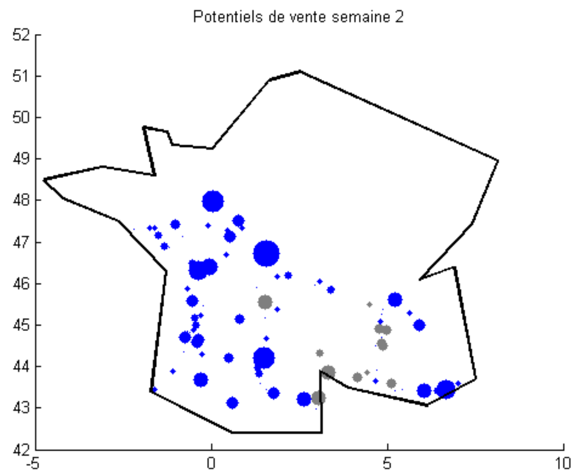

FIGURE 4 – Potentiel de vente en semaine 2, en gris les sites visités en semaine 1.

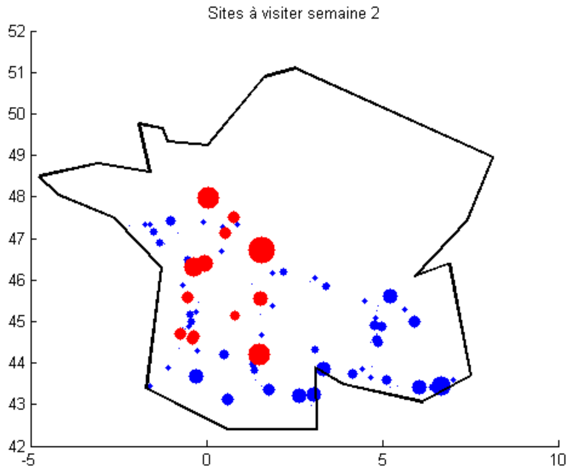

FIGURE 5 – Les sites choisis (en rouge) pour la seconde semaine.

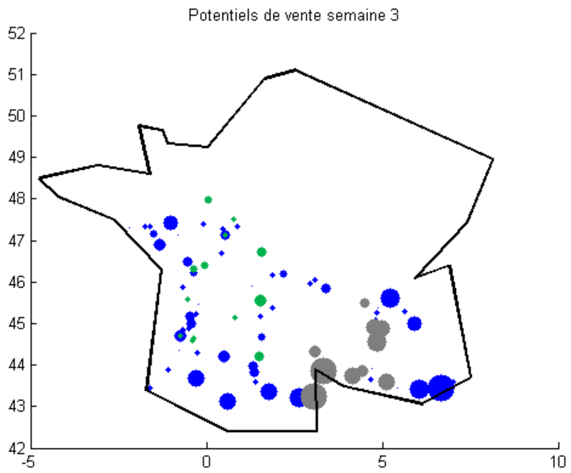

FIGURE 6 – Potentiel de vente en semaine 3, en gris ceux visités en semaine 1 et en vert ceux de la semaine 2.

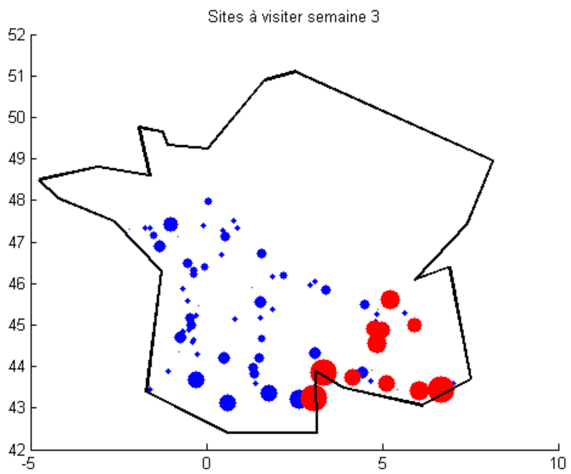

FIGURE 7 – Les sites choisis lors de la troisième semaine.

4 Conclusion

4.1 Le problème d'Articque

A partir de données de base sur un ensemble de sites (localisation géographique et chiffre d'affaire annuel), nous avons donc mis en place une modélisation d'un intérêt pour chaque point de vente. Plus les données seront variées (chiffre d'affaire saisonnier, capacité de stockage, type de produits...) et plus il sera possible de préciser cet intérêt. De plus, l'introduction d'une fonction fréquence présentée en section 2.2 permet de moduler ce facteur selon la politique commerciale de l'entreprise. Ces potentiels d'intérêt sont contrebalancés par un coût en temps lors de la tournée. Là encore plus les renseignements des temps de trajet entre deux sites seront détaillés, et plus l'optimisation proposée sera précise. Cependant nous avons pu voir que même une implémentation naïve du problème, avec des moyens calculatoires limités, a permis d'obtenir des résultats très cohérents. La rapidité d'exécution qui en résulte permet tout à fait d'envisager un développement pour un calcul en temps réel d'une feuille de route pour le visiteur.

4.2 La SEME

Cette semaine d'étude maths-entreprises nous a tout d'abord permis d'entrer dans le monde des entreprises et des industries pour plus de pratique d'une part mais elle nous a aussi permis de faire un lien entre nos études théoriques scientifiques (Mathématiques) et le monde des industries et des entreprises. De la modélisation à l'optimisation en passant par de la recherche opérationnelle, ce projet nous a permis de comprendre et de saisir toute la complexité tant théorique que pratique que ce type de problème revêt. La résolution des problèmes de ce type nécessite à la fois un bagage mathématique théorique relativement conséquent mais aussi de bonnes connaissances algorithmiques et informatiques.

Cependant il faut admettre que nous avons rencontré quelques difficultés dont les deux principales sont :

- le problème lié au temps : Résoudre un problème d'une telle ampleur en si peu de temps (4 jours) était pour nous une mission presque impossible. On a tout de suite mis une stratégie en place pour pouvoir y arriver. Cette stratégie a été pour nous la simplification du problème avec un choix de modèle assez facile à mettre en place informatiquement.

- le problème lié aux données : Disposant d'un jeu de données peu complet (fourni par Articque) avec des variables manquantes, il a été assez naturel de résoudre ce problème en faisant des suppositions et en imposant des conditions sur certaines variables ou juste en enlevant certaines contraintes à cause des variables manquantes qui y apparaissent.

Les enseignements de la SEME ont été très nombreux et riches tant sur le plan de la connaissance mais aussi sur le plan humain car la SEME nous a permis de faire la connaissance de certaines entreprises (responsables, chefs d'entreprise, ...) et de rencontrer d'autres scientifiques (chercheurs, professeurs, enseignant-chercheurs, ...).

Références

[1] Ravindra K. Ahuja, Thomas L. Magnati, and James B. Orlin. *Network flows ; Theory, Algorithms and Applications.* Prentice Hall, 1993.

[2] Enrique Benavent, Angel Corberan, José M. Sanchis, and Isaac Plana. Min-max k-vehicles windy rural postman problem. *Networks - Route 2007*, 54 :216–226, 2009.

[3] Martin Grötschel and Ya-Xiang Yuan. Euler, mei-ko kwan, königsberg, and a chinese postman. *Documenta Mathematica*, 43-50, 2012.

[4] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization ; 2nd edition.* Springer, 2006.

[5] Wilf H. S. *Algorithms and Complexity.* AK Peters, Ltd., 1994.